AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 101:333–343 (1996)

Interspecific Variation at the Y-Linked *RPS4Y* Locus in Hominoids: Implications for Phylogeny

PAUL B. SAMOLLOW, LYNN M. CHERRY, SHELLY M. WITTE, AND JEFFREY ROGERS
Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas 78245-0549

KEY WORDS Y-chromosome, Molecular evolution, Phylogeny, Humans, Apes

ABSTRACT Within- and between-species variation in restriction endonuclease recognition sites was examined at the Y-linked *RPS4Y* locus of six hominoid species: human (*Homo sapiens*), gorilla (*Gorilla gorilla*), chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), orangutan (*Pongo pygmaeus*), and gibbon (*Hylobates lar*). *RPS4Y* is an expressed gene that maps to the nonrecombining region of the Y chromosome. An approximately 1,490 base pair fragment of the *RPS4Y* gene, including all of intron 3, was amplified by PCR from DNA extracted from each of the six species. Forty-seven restriction sites were identified on the six-species composite map derived from double-digest restriction analyses of the amplified fragment. As expected, maximum parsimony analysis indicated that chimpanzee and bonobo are the two most closely related living hominoids. The same analysis suggested that the closest living relative of *Homo* is *Gorilla,* not *Pan,* although support for this relationship was relatively weak. These results disagree with recently published phylogenies based on analyses of mtDNA sequences (Horai et al. [1995] Proc. Natl. Acad. Sci. U.S.A. *88:*7401–7404) and the Y-linked *ZFY* locus (Dorit et al. [1995] Science *268:*1183–1185). A combined data set derived from three distinct Y-linked loci—*RPS4Y, SRY,* and *ZFY*—was also analyzed. The maximum parsimony topology for the combined data provided only weak support for a shared common ancestor for *Homo* and *Pan* subsequent to divergence from the *Gorilla* lineage. Taken together, the data from the Y chromosome do not provide unequivocal support for any single, dichotomously branching species tree linking *Homo, Pan,* and *Gorilla.* © 1996 Wiley-Liss, Inc.

The application of molecular methods to systematic biology has had a profound impact on our understanding of historical patterns in biotic evolution. The examination of DNA sequence differences among taxa has provided new and reliable information concerning phylogenetic relationships among a broad range of extant, and even some extinct, species. However, in some clades standard molecular systematic practice has failed to resolve problematic issues. In the case of African hominoids, comparisons of DNA sequences among the relevant genera have not yielded a clear picture of their evolutionary relationships. Data from several nuclear gene loci indicate that chimpanzees and bonobos (genus *Pan*) are more closely related to humans (genus *Homo*) than to gorillas (genus *Gorilla*); e.g., β-globin (Bailey et al., 1992), α-1,3-galactosyltransferase (Galili and Swanson, 1991), Y-chromosome pseudoautosomal boundary (Ellis et al.,

Received October 17, 1995; accepted June 16, 1996.

Address reprint requests to Paul B. Samollow, Department of Genetics, Southwest Foundation for Biomedical Research, P. O. Box 760549, San Antonio, TX 78245-0549.

© 1996 WILEY-LISS, INC.

1990). Other loci indicate that *Pan* is more closely related to *Gorilla* than to *Homo;* e.g., involucrin (Djian and Green, 1989), dopamine D4 receptor (Livak et al., 1995), and X-chromosome pseudoautosomal boundary (Ellis et al., 1990). Comparison of mitochondrial DNA (mtDNA) sequences strongly supports the former phylogenetic tree (Horai et al., 1995).

Each of the studies listed above leads to conclusions regarding the topologies of phylogenetic trees for individual loci; i.e., gene trees. However, in the context of the African hominoids, where gene trees are discordant, the most reliable estimate of species-level phylogeny will be obtained from the combined analysis of a substantial number of independent, unlinked loci (Pamilo and Nei, 1988; Rogers, 1993). The Y chromosome is a component of the mammalian genome that has not been broadly utilized for this type of phylogenetic analysis.

Except for a small pseudoautosomal region, the mammalian Y chromosome is inherited clonally and exclusively through the male lineage. Y-linked sequences should, therefore, be useful in delineating phylogenetic histories in a manner similar to, but independent of, those of mtDNA. Despite their parallel evolution, Y-specific and mtDNA histories may differ for a variety of reasons. Two of these are sex-specific differences in migration (dispersal) patterns and differences in the effective population sizes of males and females.

We have devised a PCR-based approach for the study of a Y-chromosome-specific sequence that is applicable to a broad range of primate species. Our strategy was to design PCR primers that would amplify an intron-bearing fragment within an evolutionarily conservative, single-copy Y-linked gene. We chose this approach because introns and other non-coding regions tend to evolve more rapidly than coding sequences (exons), which are selectively constrained. Intronic sequences are, therefore, more likely to provide phylogenetic information among closely related species.

The human *RPS4Y* (ribosomal protein S4, Y-linked) gene was selected as our target sequence. In humans *RPS4* comprises two closely related genes that encode isoforms of ribosomal protein S4. *RPS4Y* is located on the Y-chromosome at Yp11.3; its X-linked homolog, *RPS4X,* is located at Xq13.1 (Fisher et al., 1990). Although the two genes apparently arose from a common ancestral sequence, they lie outside of the pseudoautosomal regions and do not participate in recombinational exchange. Thus, they presumably have long, independent evolutionary histories. Human *RPS4X* and *RPS4Y* exhibit 82% cDNA base identity, and the respective 263 amino acid proteins are predicted to be 93% identical. The RPS4X and RPS4Y isoforms are both found in male ribosomes and are functionally interchangeable (Watanabe et al., 1993). Despite divergence between the X- and Y-linked homologs within a species, it appears that the *RPS4X* and *RPS4Y* sequences have been very strongly conserved among species. For example, comparison of rat (Fisher et al., 1990) and human (Wiles et al., 1988) amino acid sequences reveals 100% identity between rat RPS4 (X-linked; no Y-linked homolog has been identified in rodents) and human RPS4X proteins.

In view of these characteristics, we believe that *RPS4Y* has strong potential as a male-specific DNA sequence that can be analysed to obtain phylogenetic information which is independent of that contained in mtDNA and diploid nuclear genes. It should be useful both as an independent genetic locus for phylogenetic analysis, and as a complement to female-specific evolutionary history as revealed by mtDNA evolution. In this communication, we describe the results of an interspecific comparison of intron 3 of the *RPS4Y* gene among the living hominoids and compare these findings to phylogenetic inferences drawn from two other Y-linked genes, *SRY* (Whitfield et al., 1993) and *ZFY* (Dorit et al., 1995), and from mtDNA (Horai et al., 1995).

MATERIALS AND METHODS

DNA samples

Human high molecular weight genomic DNA was derived from lymphocytes using a modification of standard phenol/chloroform extraction procedures (Rogers and Kidd, 1993). Gorilla (*Gorilla gorilla*), chimpanzee

(*Pan troglodytes*), bonobo (*Pan paniscus*), orangutan (*Pongo pygmaeus*), and gibbon (*Hylobates lar*) DNA samples, which were extracted from lymphoblastoid cell lines, were generously provided by Amos Deinard and Kenneth Kidd (Yale University).

Molecular methods

PCR primers were designed to match exactly the exon sequences flanking intron 3 of the human *RPS4Y* gene (Fisher et al., 1990; Zinn et al., 1994). The primers were chosen to contain mismatches to the corresponding *RPS4X* sequence at their 3′ ends and consequently do not amplify the X-linked gene. The primer sequences (upstream) 5′-TTGATGGCAAGGTTCGAGTG-3′ and (downstream) 5′-CATAGACCAGGC-GGAAATGT-3′ complement sequences at the 3′ end of exon 3 and 5′ end of exon 4, respectively.

Routine PCR amplification of *RPS4Y* fragments was conducted in 100 μl reactions composed of PCR buffer (50 mM KCl, 10 mM Tris-HCl, pH 9.0, 0.1% Triton X-100 in water), 3 mM $MgCl_2$, 200 μM dGTP, dATP, dTTP, and dCTP, 0.2 μM of each of the primers, 2.5 units of *Taq* DNA polymerase (Promega), and 0.015% deionized formamide. Amplification was achieved using a Perkin/Elmer 480 thermal cycler under the following cycle conditions: pre-incubation at 95°C for 2 min, 34 cycles of 95°C for 1 min, 59°C for 1.5 min, 72°C for 2 min, 1 cycle of 95°C for 1 min, 59°C for 1.5 min, 72°C for 7 min.

Under these conditions two male-specific DNA fragments were amplified; an approximately 1,490 bp product that was amplified in all species, and a second product that varied among species (see Results). For the present study, we limited interspecific comparisons to the constant 1,490 bp fragment, which we identified as *RPS4Y* by sequence analysis. Specific amplification of this fragment was achieved by increasing the annealing temperature from 59°C (standard stringency) to the following levels (high stringency): human, gorilla, chimpanzee, bonobo, 66°C; orangutan, gibbon, 64°C. Under these conditions, only the 1,490 bp fragment was produced.

Sequence analysis of the 5′ and 3′ ends of the human 1,490 bp fragment was performed by the chain-termination method (Sanger et al., 1977) using the Sequenase™ Version 2.0 DNA sequencing kit (Amersham). The resultant sequence was compared with published human *RPS4Y* sequences (Fisher et al., 1990; Zinn et al., 1994) to verify the identity of the amplified fragment.

For single and double restriction endonuclease analyses, 20–30 μl of the reaction mixture, containing approximately 100 ng of 1,490 bp PCR product, was digested according to manufacturer's instructions. Total reaction volumes were 60 μl. Resultant digestion fragments were separated by electrophoresis on 1% agarose gels or 8% polyacrylamide gels (depending on fragment sizes) and visualized by ethidium bromide staining.

Phylogenetic analysis

The restriction site maps for the *RPS4Y* sequence of the six hominoid species were analysed using maximum parsimony by means of the computer program MacClade (Maddison and Maddison, 1992). The gibbon and orangutan maps were both defined as the outgroups. All restriction sites observed to occur in any one species were treated as characters with two alternative states (present or absent). Two types of analyses were performed based on assumptions about the likelihood of mutational gains and losses of restriction sites: 1) assuming that the gain of a site and the loss of a site are equally likely events, all interspecies changes in character states were weighted equally; 2) under the alternative assumption of Dollo parsimony, as implemented via MacClade, a gain of a site was allowed to occur only one time within the phylogenetic tree. Within a small set of closely related species, this restriction is equivalent to heavy weighing of the relative probabilities of character state changes such that site gains are much less likely than site losses.

Following the analysis of *RPS4Y* alone, we combined data from three Y-linked loci: *RPS4Y, SRY,* and *ZFY.* The *RPS4Y* analyses were based on restriction site data. The *SRY* and *ZFY* analyses were based on 205 bp and 729 bp of nucleotide sequence data (Whitfield et al., 1993; Dorit et al., 1995), respec-

tively. The single, combined data set was analyzed using maximum parsimony as described above. No gibbon sequence is available for *ZFY,* so all the character states for this locus were treated as unknown, leaving the orangutan sequence as the only outgroup.

RESULTS

Amplification of *RPS4Y*

The PCR primers described amplified a male-specific product (Fig. 1) of 1,490 bp from DNA of all primate species tested: the six hominoids (human, gorilla, chimpanzee, bonobo, orangutan, gibbon), as well as baboon (*Papio hamadryas*), rhesus monkey (*Macaca mulatta*), drill (*Mandrillus leucophaeus*)/mandrill (*M. sphinx*) hybrids, and two species of squirrel monkeys (*Saimiri boliviensis* and *S. sciurius*) (Samollow et al., 1994). A second male-specific fragment, varying in size from approximately 1,300 bp to approximately 1,700 bp, depending on species, was also amplified using the same primers (Samollow et al., 1994) under standard stringency, but was eliminated by high stringency conditions. Attempts to amplify DNA obtained from females of these same 11 species (combined N = 48) using these primers failed to produce either fragment under a broad range of conditions (Fig. 1). A preliminary analysis of restriction digestion patterns indicates that the 1,300/1,700 bp fragments are distinct from the 1,490 bp product. They have not been characterized further.

Sequence data were obtained for 119 bp of the 5′ end and 78 bp of the 3′ end of the 1,490 bp human product amplified under high stringency conditions (data not shown). Disregarding the PCR primer sequences, the fragment contained 28 bases of coding sequence at each end. Together with the additional 20 bases of intron 3 sequence published by Zinn et al. (1994), we were able to compare a total of 76 bp with published data (Fisher et al., 1990). The 76 bp of our fragment matched the published human *RPS4Y* sequences exactly.

Restriction endonuclease site variation

RPS4Y intron 3 PCR products from four humans, two gorillas, two chimpanzees, three bonobos, one orangutan, and one gibbon were subjected to digestion by 35 restriction endonucleases. Of these, 16 cut the fragment at one or more sites in at least one of the six species (Table 1). A total of 47 restriction sites were detected among the six species. After determining the presence or absence of polymorphism (discussed below) using the 16 relevant enzymes, a single sample from each species was subjected to a series of double digestions to generate a species restriction map. Figure 2 shows a composite restriction map summarizing all of the 47 restriction sites detected in the six species combined.

Inspection of Figure 2 reveals 28 sites whose positions on the restriction map could not be distinguished from those of one or more other sites. These represent potentially overlapping sequences. Where restriction sites actually overlap, there is potential for a single base change to alter the states of two or more different characters (sites), with the possible effect of biasing the inferred phylogeny. We examined this issue by comparing restriction endonuclease recognition sequences among potentially overlapping site pairs, trios, and quartets. In doing so we assumed that 1) each change in a character state among species was due to a single mutational change, and 2) derived character states among species are homologous; i.e., the presence or absence of a restriction site in any one species is due to the same base sequence as the presence or absence of that same site in any other species. Based on these criteria, 17 of the 28 clustered sites could be eliminated as overlaps. Of the remaining 11 sites (sites 15/16, 19/20, 37/38/40, 41/42, and 44/45), four (15/16 and 41/42) were present in all species and thus were of no phylogenetic significance. Of the remaining seven sites, five (37/38/40 and 44/45) varied only between the two outgroup species and two (19/20) represented chimpanzee/bonobo synapomorphies (*Pan* autapomorphies). Thus, none of these seven potentially overlapping sites had any bearing on the question of relationships among *Homo, Pan,* and *Gorilla.* Elimination of any (or all) of these sites from the analysis reduces the total length of the maximum parsimony tree or alters the number of changes

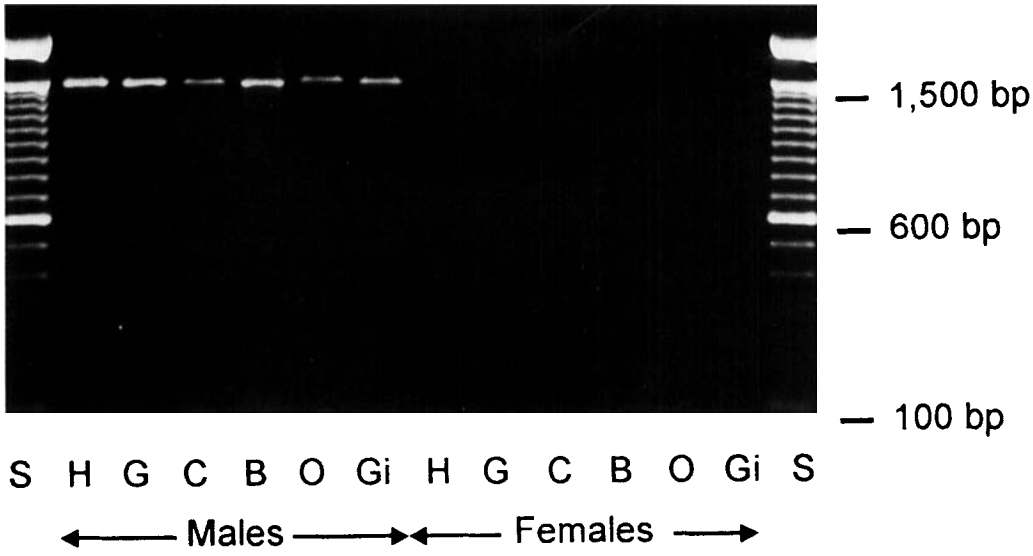

Fig. 1. *RPS4Y* intron 3 amplification from African hominoid DNA. Fragments were amplified from genomic DNA extracts under high stringency conditions (see text). Samples: S = 100 bp ladder standard; H = human; G = gorilla; C = chimpanzee; B = bonobo; O = orangutan; Gi = gibbon.

between the gibbon and orangutan branches, but has no influence on the branching topology within the *Homo–Pan–Gorilla* clade. For this reason, we present the analysis of the data with all 47 sites retained.

Considerable interspecific variation in the presence of restriction sites was observed among the six species (Table 1). The number of sites present in a given species ranged from 28 in gibbon to 33 in human, chimpanzee, and bonobo. Only 18 sites were present in all species examined (Table 1 and Fig. 2). Variation was also detected in the length of the *RPS4Y* PCR product. Our estimate of the total length of the amplification product is approximately 1,490 bp for all species examined except orangutan and bonobo. The orangutan product was consistently larger due to an apparent 10–15 bp insertion in the region between restriction sites 25 and 32. Similarly, the total length of the bonobo fragment appeared slightly smaller than that of human, gorilla, chimpanzee, and gibbon due to an apparent 5–10 bp deletion between sites 4 and 7. We have not considered these two minor size differences as characters in our analysis of phylogenetic relatedness.

Both of the chimpanzees and two of the three bonobos exhibited identical restriction digestion patterns at all sites. The third bonobo lacked site 18, a *Hin*fI site. Replicate amplification of these samples verified that the variation among bonobos was genuine rather than a PCR amplification artifact (data not shown). No other intraspecific variation was detected in this study.

Phylogenetic analysis

Analysis of the full 47 character *RPS4Y* restriction site data set using unweighted character states yielded a maximum parsimony tree that places *Homo* and *Gorilla* as sister taxa. However, support for this branching scheme is weak; it requires 33 mutational steps, while the trees linking *Pan* and *Homo,* or *Pan* and *Gorilla,* as sister taxa require only one additional step (Table 2). We also performed phylogenetic analyses for *RPS4Y* using lesser data sets produced by elimination of potentially overlapping sites. In every case, the topologies of the resulting phylogenetic trees were identical to those using the full data set. Only the total length of the trees changed, due to reduction in the number of mutational differences

TABLE 1. Restriction endonuclease recognition sites in RPS4Y *intron 3 fragment of humans and apes*

Site no.	Restriction endonuclease	Human	Gorilla	Chimpanzee	Bonobo	Orangutan	Gibbon
1	*Hin*fI	+	+	+	+	+	+
2	*Taq*I	−	−	+	+	−	−
3	*Ban*II	+	−	+	+	+	+
4	*Hpa*II	−	−	+	+	+	+
5	*Pst*I	+	−	−	−	+	−
6	*Ava*II	+	+	+	+	+	+
7	*Hae*III	+	+	+	+	+	+
8	*Hpa*II	+	−	−	−	+	−
9	*Nci*I	+	−	−	−	−	−
10	*Rsa*I	−	+	−	−	−	−
11	*Dde*I	+	+	+	+	+	+
12	*Hind*III	−	−	−	−	+	+
13	*Ban*II	+	+	+	+	+	+
14	*Xba*I	−	+	−	−	−	−
15	*Dde*I	+	+	+	+	+	+
16	*Hind*III	+	+	+	+	+	+
17	*Ban*II	+	+	+	+	+	+
18	*Hin*fI	+	+	+	+/−[1]	+	+
19	*Hpa*II	−	−	+	+	−	−
20	*Nci*I	−	−	+	+	−	−
21	*Sty*I	+	+	+	+	+	+
22	*Dde*I	−	−	−	−	+	−
23	*Rsa*I	+	+	+	+	+	+
24	*Dde*I	−	−	−	−	−	+
25	*Eco*RI	+	+	+	+	+	−
26	*Dde*I	+	−	−	−	−	−
27	*Rsa*I	+	−	−	−	−	−
28	*Hae*III	+	−	+	+	+	−
29	*Sty*I	+	+	+	+	+	−
30	*Hin*fI	−	−	−	−	−	+
31	*Hpa*II	+	+	+	+	−	−
32	*Bam*HI	+	+	+	+	+	+
33	*Dde*I	+	+	+	+	+	+
34	*Hin*fI	+	+	+	+	+	−
35	*Dde*I	+	+	+	+	−	+
36	*Hae*III	+	+	+	+	+	+
37	*Hin*fI	−	−	−	−	−	+
38	*Sty*I	−	−	−	−	+	−
39	*Dde*I	+	+	+	+	−	+
40	*Hae*III	−	−	−	−	+	−
41	*Bam*HI	+	+	+	+	+	+
42	*Dde*I	+	+	+	+	+	+
43	*Cfo*I	−	−	+	+	−	+
44	*Hin*fI	+	+	+	+	+	+
45	*Sty*I	+	+	+	+	+	−
46	*Dde*I	+	+	+	+	+	+
47	*Taq*I	+	+	+	+	−	+

[1] Intraspecific variation: see text.

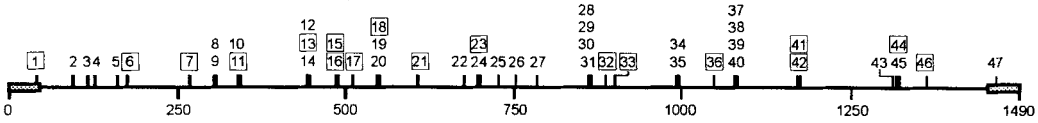

Fig. 2. Composite restriction site map of *RPS4Y* intron 3 fragment and flanking exon sequences. Line indicates intron 3 region; shaded boxes represent exon fragments. Numbered restriction sites appearing above the fragment correspond to sites listed in Table 1. Boxed sites are present in all species examined (human, gorilla, chimpanzee, bonobo, orangutan, gibbon). The relative positions of sites that are shown stacked upon one another could not be distinguished. Numbers below the fragment indicate approximate base pairs from the 5′ end of the fragment (O = transcript position 215, within exon 3).

TABLE 2. Maximum parsimony analysis of RPS4Y, SRY, *and* ZFY *data from humans and apes*

Loci included	Sister taxa[1]		
	Homo–Pan	*Homo–Gorilla*	*Pan–Gorilla*
RPS4Y + SRY + ZFY[2]	117	119	120
RPS4Y + SRY[2]	78	77	78
RPS4Y[2]	34	33	34
SRY[2]	44	44	44
ZFY[3]	39	42	42

[1]Number indicates the minimum number of mutational steps required to force the indicated sister taxon pair using the data from the loci included in the analysis.
[2]Outgroups = gibbon and orangutan.
[3]Outgroup = orangutan only.

among the outgroup species and/or among chimpanzee and bonobo. Analyses of these reduced data sets are not considered in this report.

Application of character state weighing via the Dollo restriction of MacClade (see above) resulted in two equally parsimonious trees: *Homo–Gorilla* and *Pan–Gorilla.* Each of these trees required 35 mutational steps. Under this scheme, the *Homo–Pan* solution was the least parsimonious of the three dichotomously branching solutions, requiring two additional steps (37 total steps). Interestingly, this is the same number of steps (37) required by the solution in which *Homo, Pan,* and *Gorilla* diverge trichotomously.

Analysis of the *SRY* and *ZFY* data (both unweighted) yields very different results. The *SRY* sequences show no characters that link any particular pair in the *Pan, Homo, Gorilla* trio more closely than any other pair (Table 2). Thus, *SRY* suggests a trichotomous divergence of the three lineages. *ZFY* data support the *Homo–Pan* clade (Dorit et al., 1995), although support for this conclusion also is weak (Rogers et al., 1996). Since the Y chromosome is inherited as a single, non-recombining unit, the data from these three loci can be combined into a single haplotype. In the combined *RPS4Y–SRY–ZFY* tree (Fig. 3) there are three synapomorphies uniquely linking *Homo* and *Pan* to the exclusion of *Gorilla.* All of these synapomorphies are in the *ZFY* sequence; there are no synapomorphies in the *RPS4Y* or *SRY* sequences. This contrasts with the terminal branches leading to *Homo, Pan,* and *Gorilla,* where *ZFY* exhibits fewer changes (10 changes) than *RPS4Y* (13 changes) or *SRY* (12 changes). The maximum parsimony *Homo–Pan* tree for the combined data set is two steps shorter (117 vs. 119) than the next most parsimonious one, which links *Homo* and *Gorilla.* However, if *ZFY* is eliminated from the analysis (discussed below), the combined *RPS4Y–SRY* data weakly support the *Homo–Gorilla* topology.

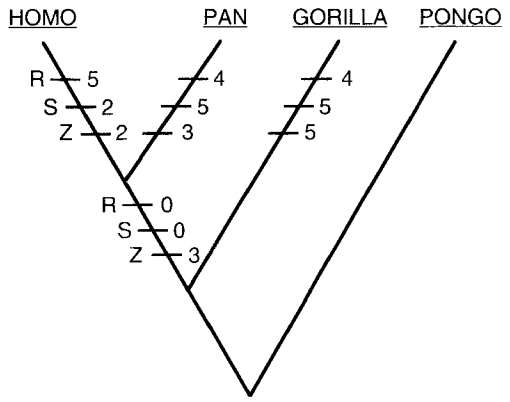

Fig. 3. Maximum parsimony tree for 981 characters from the combined *RPS4Y–SRY–ZFY* data set in African hominoids. The tree is rooted using orangutan as the outgroup. Numbers adjacent to cross hatches indicate the number of unambiguous mutational changes along the branch for the gene indicated: R = *RPS4Y;* S = *SRY;* Z = *ZFY.*

DISCUSSION

We have utilized a PCR-amplified fragment of a single-copy, Y-linked gene, *RPS4Y,* to examine evolutionary relationships among living hominoid species. Several features of the *RPS4Y* gene indicate that it has strong potential as a male-specific character for the examination of primate population structure and phylogenetic relationships. The approximately 1,490 bp, intron-bearing fragment can be amplified from a broad range

of primates including hominoids, Old World monkeys, and New World monkeys (e.g., *Saimiri*) using one set of PCR primers and virtually identical reaction conditions. Yet, the fact that all 11 species examined to date can be discriminated on the basis of a few restriction digest patterns suggests that the intron sequences of this gene evolve rapidly enough to provide clear evolutionary signal.

The present analysis reveals that the *RPS4Y* sequence of humans is more closely related to that of gorillas than to that of chimpanzee or bonobos, a finding that is contrary to the majority of molecular data available for this clade (e.g., Rogers, 1993; Bailey, 1993; Horai et al., 1995). However, this relationship is based on the existence of only a single human–gorilla synapomorphy. This level of support is so weak that it cannot be regarded as compelling evidence concerning evolutionary relationships among hominoid Y chromosomes. In our view, the phylogenetic implications of the *RPS4Y* data are best interpreted in a context that considers data available from additional Y-linked sequences.

The inheritance of the Y chromosome as a single, non-recombining unit implies that all (non-pseudoautosomal) segments of the Y chromosome should exhibit the same phylogeny for any group of species. Differences in phylogenetic patterns among Y-linked loci of closely related taxa, such as *Homo, Pan,* and *Gorilla,* could not arise as the result of random fixations of alleles in a polymorphic series that existed in their last common ancestor. Thus, data derived from Y-chromosome sequences are different from similar data derived from autosomal or X-linked loci, because all discrepancies among Y-linked loci must be attributed to convergent- or back-mutations, not ancient polymorphism. In this light, we note that three Y-chromosomal regions, *RPS4Y, SRY,* and *ZFY,* do not yield the same gene tree topology for *Homo, Pan,* and *Gorilla. RPS4Y* suggests that the closest relative of *Homo* is *Gorilla,* while *ZFY* suggests it is *Pan.* The *SRY* locus does not support either of these dichotomously branching trees. These discrepancies force us to conclude that some homoplasy exists in the overall Y-chromosome data set, and invite closer examination of the data.

In a recent analysis of Y-chromosome evolution, Dorit et al. (1995) reported a maximum parsimony reconstruction of a 729 bp region of the *ZFY* gene which indicated that chimpanzee/bonobo *ZFY* sequence is more closely related to human *ZFY* than to the homologous sequence in gorillas. If this reconstruction accurately reflects species-level phylogenetic history, it would date the divergence of *Gorilla* approximately 2.5 million years prior to the divergence of *Homo* from *Pan* (Dorit et al., 1995; Rogers et al., 1996). However, two features of the *ZFY* sequence cast doubt on the correspondence between the *ZFY* gene-level phylogeny proposed by Dorit et al. (1995) and the actual history of diversification among these evolutionary lineages.

First, *ZFY* genes in several mammals, including primates, appear to have undergone gene conversion (Hayashida et al., 1992; Pamilo and Bianchi, 1993). Gene conversion can obscure phylogenetic information in molecular data. Our present maximum parsimony solution using the combined *RPS4Y, SRY,* and *ZFY* data set provides weak support for the *Homo–Pan* relationship. However, if out of caution over the issue of gene conversion we exclude the *ZFY* data from the present analyses, then the preferred phylogenetic tree for the Y chromosome links *Homo* and *Gorilla* (requiring 77 mutational steps) rather than *Homo* and *Pan* (78 steps). Second, we have recently argued (Rogers et al., 1996) that the phylogenetic topology and internodal length suggested by the *ZFY* data are paradoxical when contrasted to established paleontological evidence concerning the origin of hominids and the divergence date for *Pan* and *Homo.* We show that this paradox can be avoided by adopting the second most parsimonious solution for the *ZFY* data: a trichotomous divergence that requires only two more steps (72 rather than 70) than the most parsimonious, dichotomously branching tree. If this second most parsimonious reconstruction is substituted for the most parsimonious one in the combined *RPS4Y–SRY–ZFY* tree, support for any sort of dichotomous branching scheme virtually disappears. Consequently, the Y-chromosome data currently available do not resolve unambiguously the phylogenetic re-

lationships among the three lineages leading to *Homo, Pan,* and *Gorilla.*

The idea of a trichotomous branching scheme for the Y chromosome of the African hominoid clade is at odds with conclusions based on analyses of mtDNA sequences (Horai et al., 1995) and some diploid nuclear genes (Takahata, 1995), which yield estimates of the *Gorilla–*[*Homo–Pan*] internodal period of 1.7 million years and 2.6 million years, respectively. The number of *Homo–Pan* synapomorphies expected in the relevant internodal period can be determined by calculating the average rate of Y-chromosome evolution using the observed number of mutations that have occurred along each of the terminal branches of the phylogenetic tree (see Fig. 3). We used the two estimates of the length of the *Gorilla–*[*Homo–Pan*] internode mentioned above (1.7 and 2.6 million years) to estimate the expected number of *Homo–Pan* synapomorphies for the combined *RPS4Y–SRY–ZFY* data set. Based on these estimates we would expect between 3.6 and 5.6 mutational changes along a *Gorilla–*[*Homo–Pan*] internode 1.7 to 2.6 millions years in length. We observed three such changes in our combined maximum parsimony tree, all of which were in the *ZFY* sequence. If we utilize the second most parsimonious reconstruction for *ZFY* (i.e., the trichotomous tree), then we would conclude that no mutational changes occurred during the putative internodal period. If, more conservatively, we simply exclude *ZFY* from consideration, we still expect 2.6 to 4.0 changes among *RPS4Y* and *SRY* during the putative internode, but none were observed in the combined *RPS4Y–SRY* data set. Thus, depending on how the *ZFY* data are viewed, and what phylogenetic topology is adopted for this locus, the observed number of Y-chromosomal synapomorphies in the putative *Gorilla–*[*Homo–Pan*] internode may suggest a shorter time span than those inferred from mtDNA and diploid nuclear gene data sets.

The concept of a trichotomous divergence of the Y chromosome is consistent with two alternative ideas concerning the *Homo–Pan–Gorilla* radiation. One possibility is that the mtDNA data are in error, and the conclusions of Horai et al. (1995) regarding the evolution of hominoid mtDNA are incorrect. This seems very unlikely. The mtDNA data are strong and appear to establish firmly that the mtDNA of *Pan* is more similar to that of *Homo* than to that of *Gorilla.* This suggests the second, alternative interpretation, that the Y-chromosome and mtDNA phylogenies may actually differ, either in overall topology or absolute length of internodal period. This interpretation proposes that the 1.7 million year internode for the mtDNA gene tree is correct, but that it does not accurately reflect the timing of the overall genetic isolation and subsequent divergence of the *Gorilla* and *Homo–Pan* lineages. If this is true, a possible explanation could be that differences in sex-specific patterns of dispersal among populations of the last common ancestor of the three lineages led to differences in the divergence patterns for male-specific (Y-linked) and female-specific (mtDNA) sequences. Such a difference in dispersal patterns could result in a protracted period of male-specific gene flow among the incipient lineages, during which female-specific gene flow was minimal, possibly due to more pronounced philopatry. This more complex pattern of dispersal and differentiation could account for both the long internodal period implied by the mtDNA data and the much shorter one suggested by the Y-chromosome data. It would also be consistent with the disparate phylogenetic topologies that have been observed among the trees of several diploid nuclear genes of these species (Rogers, 1993). Modern hominoids exhibit a broad range of male/female dispersal patterns, and there is not presently, and may never be, a reliable means by which the dispersal pattern of the long-extinct common ancestor of these modern lineages can be determined. However, male dispersal/female philopatry is the most common pattern among the extant catarrhines that have been examined (Pusey and Packer, 1986).

While Y-linked, autosomal, and mtDNA sequences should generally yield comparable patterns of phylogenetic divergence, results from analyses of these different genomic subsets can (and do) occasionally differ. Such differences may reflect the general problems inherent in inferring species-level

phylogenetic relationships from individual loci (see Rogers, 1994), but factors related to the distinct patterns of inheritance of these sequences may also play a role. As proposed above, sex differences in migration patterns (and/or effective population sizes) could lead to distinct patterns of geographic distribution for Y chromosomes and mtDNA, and thereby result in the differential distribution of variation in Y-linked and mtDNA sequences among subdivided populations. Under some circumstances such population-level differentiation could eventually lead to species-level differences in the distribution of Y-chromosomal and mtDNA variation. We observed intraspecific variation at the *RPS4Y* locus of *Pan paniscus* in the present study. Such Y-linked intraspecific variants would be valuable tools in studies of sex-specific patterns of dispersal, population subdivision, and hybridization. Investigation of *RPS4Y* in a variety of primate species will likely produce valuable information regarding species-level phylogeny and may be useful in the examination of intraspecific patterns of genetic differentiation as well.

ACKNOWLEDGMENTS

We thank Anthony G. Comuzzie for thoughtful discussions of many of the issues considered herein, and Amos Deinard and Kenneth K. Kidd for the gorilla, chimpanzee, bonobo, orangutan, and gibbon DNA samples used in this study.

LITERATURE CITED

Bailey WJ, Hayasaka K, Skinner CG, Kehoe S, Sieu LC, Slightom JL, and Goodman M (1992) Reexamination of the African hominoid trichotomy with additional sequences from the primate β-globin gene cluster. Mol. Phylogenet. Evol *1*:97–135.

Bailey WJ (1993) Hominoid trichotomy: A molecular overview. Evol. Anthropol. *2*:100–108.

Djian P, and Green H (1989) Vectorial expansion of the involucrin gene and the relatedness of the hominoids. Proc. Natl. Acad. Sci. U.S.A. *86*:8447–8451.

Dorit RL, Akashi H, and Gilbert W (1995) Absence of polymorphism at the *ZFY* locus on the human Y chromosome. Science *268*:1183–1185.

Ellis N, Yen P, Neiswanger K, Shapiro LJ, and Goodfellow PN (1990) Evolution of the pseudoautosomal boundary in the Old World monkeys and great apes. Cell *63*:977–986.

Fisher EMC, Beer-Romero P, Brown LG, McNeil JA, Lawrence JB, Willard HF, Bieber FR, and Page DC (1990) Homologous ribosomal protein genes on the human X and Y chromosomes: Escape from X inactivation and possible implications for Turner syndrome. Cell *63*:1205–1218.

Galili U, and Swanson K (1991) Gene sequences suggest inactivation of α-1,3-galactosyltransferase in catarrhines after divergence of apes from monkeys. Proc. Natl. Acad. Sci. U.S.A. *88*:7401–7404.

Hayashida H, Kuma K, and Miyata T (1992) Interchromosomal gene conversion as a possible mechanism for explaining divergence patterns of ZFY-related genes. J. Mol. Evol. *35*:181–183.

Horai S, Hayasaka K, Kondo R, Tsugane K, and Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. U.S.A. *92*:532–536.

Livak KJ, Rogers J, and Lichter JB (1995) Variability of dopamine D_4 receptor (*DRD4*) gene sequence within and among nonhuman primate species. Proc. Natl. Acad. Sci. U.S.A. *92*:427–431.

Maddison WP, and Maddison DR (1992) MacClade. Analysis of Phylogeny and Character Evolution. Version 3. Sunderland: Sinauer.

Pamilo P, and Bianchi NO (1993) Evolution of the *Zfx* and *Zfy* genes: Rates and interdependence between the genes. Mol. Biol. Evol. *10*:221–281.

Pamilo P, and Nei M (1988) Relationship between gene trees and species trees. Mol. Biol. Evol. *5*:568–583.

Pusey AE, and Packer C (1986) Dispersal and philopatry. In BB Smuts, DL Cheney, RM Seyfarth, RW Wrangham, and TT Struhsaker (eds.): Primate Societies. Chicago: University of Chicago Press, pp. 250–266.

Rogers J (1993) The phylogenetic relationships among *Homo, Pan,* and *Gorilla:* A population genetics perspective. J. Hum. Evol. *25*:201–215.

Rogers J (1994) Levels of genealogical hierarchy and the problem of hominoid phylogeny. Am. J. Phys. Anthropol. *94*:81–88.

Rogers J, and Kidd KK (1993) Nuclear DNA polymorphisms in a wild population of yellow baboons (*Papio hamadryas cynocephalus*) from Mikumi National Park, Tanzania. Am. J. Phys. Anthropol. *90*:477–486.

Rogers J, Samollow PB, and Comuzzie AG (1996) Technical comment: Estimating the age of the common ancestor of men from the *ZFY* intron. Science *272*: 1360–1361.

Samollow PB, Cherry LM, and Rogers J (1994) Sex identification and species discrimination via PCR amplification of a Y-linked ribosomal gene fragment. Isozyme Bull. *27*:50.

Sanger F, Nicklen S, and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. *74*:5463–5467.

Takahata N (1995) A genetic perspective on the origin and history of humans. Annu. Rev. Ecol. Syst. *26*:343–372.

Watanabe M, Zinn AR, Page DC, and Nishimoto T (1993) Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. Nat. Genet. *4*:268–271.

Whitfield LS, Lovell-Badge R, and Goodfellow PN (1993)

Rapid sequence evolution of the mammalian sex-determining gene *SRY*. Nature *364*:713–715.

Wiles MV, Alexander CM, and Goodfellow PN (1988) Isolation of an abundantly expressed sequence from the human X chromosome by differential screening. Somat. Cell Mol. Genet. *14*:31–39.

Zinn A, Alagappan RK, Brown LG, Wool I, and Page DC (1994) Structure and function of ribosomal protein S4 genes on the human and mouse sex chromosomes. Mol. Cell. Biol. *14*:2485–2492.